Homework 2 SOLUTIONS (written)

Problem 1 The full generative process is

$$\text{Data: } y_i \overset{iid}{\sim} \text{Bern}(\pi), \quad x_{i,d}|y_i \sim \text{Pois}(\lambda_{y_i,d}),\ d = 1, \dots, D \qquad \text{Prior: } \lambda_{y,d} \overset{iid}{\sim} \text{Gamma}(2, 1)$$

Solve

$$\widehat{\pi}, \widehat{\lambda}_{0,1:D}, \widehat{\lambda}_{1,1:D} = \arg \max_{\pi, \widehat{\lambda}_{0,1:D}, \widehat{\lambda}_{1,1:D}} \sum_{i=1}^{n} \ln p(y_i|\pi) + \sum_{d=1}^{D} \left(\ln p(\lambda_{0,d}) + \ln p(\lambda_{1,d}) + \sum_{i=1}^{n} \ln p(x_{i,d}|\lambda_{y_i,d}) \right).$$

We have that

$$p(y_i|\pi) = \pi^{y_i}(1-\pi)^{1-y_i}, \qquad p(\lambda_{y,d}) = \lambda_{y,d}\mathrm{e}^{-\lambda_{y,d}}$$

$$p(x_{i,d}|y_i) = \frac{\lambda_{y_i,d}^{x_{i,d}}}{x_{i,d}!}\mathrm{e}^{-\lambda_{y_i,d}} = \left(\frac{\lambda_{1,d}^{x_{i,d}}}{x_{i,d}!}\mathrm{e}^{-\lambda_{1,d}}\right)^{y_i}\left(\frac{\lambda_{0,d}^{x_{i,d}}}{x_{i,d}!}\mathrm{e}^{-\lambda_{0,d}}\right)^{1-y_i}$$

The objective is

$$\mathcal{L} = \sum_{i=1}^{n} \ln y_i \ln \pi + (1-y_i)\ln(1-\pi) + const.$$

$$+ \sum_{d=1}^{D} \left(\ln \lambda_{1,d} - \lambda_{1,d} + \ln \lambda_{0,d} - \lambda_{0,d} + \sum_{i=1}^{n} y_i(x_{i,d} \ln \lambda_{1,d} - \lambda_{1,d}) + (1-y_i)(x_{i,d} \ln \lambda_{0,d} - \lambda_{0,d}) \right)$$

(a)

$$\frac{\partial \mathcal{L}}{\partial \pi} = 0 = \sum_{i=1}^{n} \frac{y_i}{\pi} - \frac{1-y_i}{1-\pi} \Rightarrow \pi = \frac{1}{n}\sum_{i=1}^{n} y_i$$

(b) Taking $\lambda_{1,d}$ as example

$$\frac{\partial \mathcal{L}}{\partial \lambda_{1,d}} = 0 = \frac{1}{\lambda_{1,d}} - 1 + \sum_{i=1}^{n}\left(\frac{y_i x_{i,d}}{\lambda_{1,d}} - y_i\right) \Rightarrow \lambda_{1,d} = \frac{1 + \sum_{i=1}^{n} y_i x_{i,d}}{1 + \sum_{i=1}^{n} y_i}$$

To cover all possible cases, can equivalently use indicators for $y \in \{0, 1\}$. Results in

$$\lambda_{y,d} = \frac{1 + \sum_{i=1}^{n} 1(y_i = y)x_{i,d}}{1 + \sum_{i=1}^{n} 1(y_i = y)}$$

Problem 2

(e)

$$\mathcal{L}(w) \approx \mathcal{L}'(w) \equiv \mathcal{L}(w_t) + (w - w_t)^T \nabla \mathcal{L}(w_t) + \frac{1}{2}(w - w_t)^T \nabla^2 \mathcal{L}(w_t)(w - w_t)$$

$$\nabla_w \mathcal{L}'(w) = 0 = \nabla \mathcal{L}(w_t) + \nabla^2 \mathcal{L}(w_t) w - \nabla^2 \mathcal{L}(w_t) w_t$$

$$w = w_t - (\nabla^2 \mathcal{L}(w_t))^{-1} \nabla \mathcal{L}(w_t) \Rightarrow w_{t+1}$$

$$\mathcal{L} = \sum_{i=1}^{n} \ln \sigma(y_i x_i^T w_t), \qquad \sigma(y_i x_i^T w_t) = \frac{1}{1 + \mathrm{e}^{-y_i x_i^T w_t}}$$

$$\nabla \mathcal{L}(w_t) = \sum_{i=1}^{n} (1 - \sigma(y_i x_i^T w_t)) y_i x_i$$

$$\nabla^2 \mathcal{L}(w_t) = -\sum_{i=1}^{n} \sigma(y_i x_i^T w_t)(1 - \sigma(y_i x_i^T w_t)) x_i x_i^T$$